The Synthesizability of Molecules Proposed by Generative Models

Year: 2020 | Citations: 285 | Authors: Wenhao Gao, Connor W. Coley

---

Abstract

The discovery of functional molecules is an expensive and time-consuming process, exemplified by the rising costs of small molecule therapeutic discovery. One class of techniques of growing interest for early-stage drug discovery is de novo molecular generation and optimization, catalyzed by the development of new deep learning approaches. These techniques can suggest novel molecular structures intended to maximize a multi-objective function, e.g., suitability as a therapeutic against a particular target, without relying on brute-force exploration of a chemical space. However, the utility of these approaches is stymied by ignorance of synthesizability. To highlight the severity of this issue, we use a data-driven computer-aided synthesis planning program to quantify how often molecules proposed by state-of-the-art generative models cannot be readily synthesized. Our analysis demonstrates that there are several tasks for which these models generate unrealistic molecular structures despite performing well on popular quantitative benchmarks. Synthetic complexity heuristics can successfully bias generation toward synthetically-tractable chemical space, although doing so necessarily detracts from the primary objective. This analysis suggests that to improve the utility of these models in real discovery workflows, new algorithm development is warranted.